

- В. А. Анфилов // Межвуз. сб. науч. тр.: Проблемы экономики, финансов и управления производством. Иваново: ИГХТУ, 2015. – № 15.
3. Воробьев, Н. Н. Современное состояние теории игр [Текст] / Н. Н. Воробьев // Успехи мат, наук. — М. : Знание, 1976.
 4. Нейман, Дж. Теория игр и экономическое поведение [Текст] / Дж. Нейман, О. Моргенштерн // М.: Наука, 1970.
 5. Федурин, С. А. Игровая модель электронных аукционов [Текст]: вып. квал. работа / С. А. Федурин // Кафедра теоретической экономики, 2015.
 6. Андреев Н.Ю., Кордыш Ф.С. Использование электронных торгов в современной контрактной системе // Имущественные отношения в Российской Федерации. 2015.
 7. Храмкина А.А. О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд. / Сборник докладов. М.: ИД «Юриспруденция», 2014.

ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОИСКА НЕФТЕНАСЫЩЕННЫХ ПЛАСТОВ

Э.В. Шаякберов

Научный руководитель Трусова А.Ю.

В данной работе будет оценен потенциал применения методов машинного обучения для восстановления каротажей насыщения – кривой, которая наиболее точно может дать ответ о наличии нефти в данном интервале.

Первым этапом работы является подготовка исходных данных. Для расчётов были собраны каротажные скважины Оренбургской области. Данные

месторождения расположены близко друг к другу и имеют схожую геологию (Рисунок 1). Суммарно, 87 скважин и 2913 каротажных кривых.

В ходе подготовки данных были выбраны каротаж по скважинам месторождений Оренбургской области. Суммарно, 87 скважин и 2913 каротажных кривых. Размер данных – 1 Гб. Месторождения имеют схожее геологическое строение и расположены близко друг к другу.

В ходе обработки были убраны каротаж без потенциальной физической связи с искомым каротажом (например, DS) и убраны каротаж, имеющие высокую степень корреляции (часть GZ, MGZ/MPZ), чтобы не тратить дополнительное процессорное время и не зачумлять выборку.

Для работы были выбраны следующие каротаж: GK, PS, GZ1-GZ5, IK, BK, MBK, NGK, DTP, MGZ, MPZ, PS, PZ, REZ, которые присутствовали в большинстве скважин выборки.

Наиболее дорогостоящим, редким и достоверным является каротаж ИННК. Он с высокой точностью показывает наличие нефти на нужной нам глубине. Восстановление данного каротажа позволило бы автоматизировать нахождение нефти по различным глубинам. Данный каротаж присутствует во всех скважинах выборки. Таким образом, задача будет сведена к восстановлению каротажа ИННК.

С помощью оставшихся кривых принято решение восстанавливать кривую ИННК.

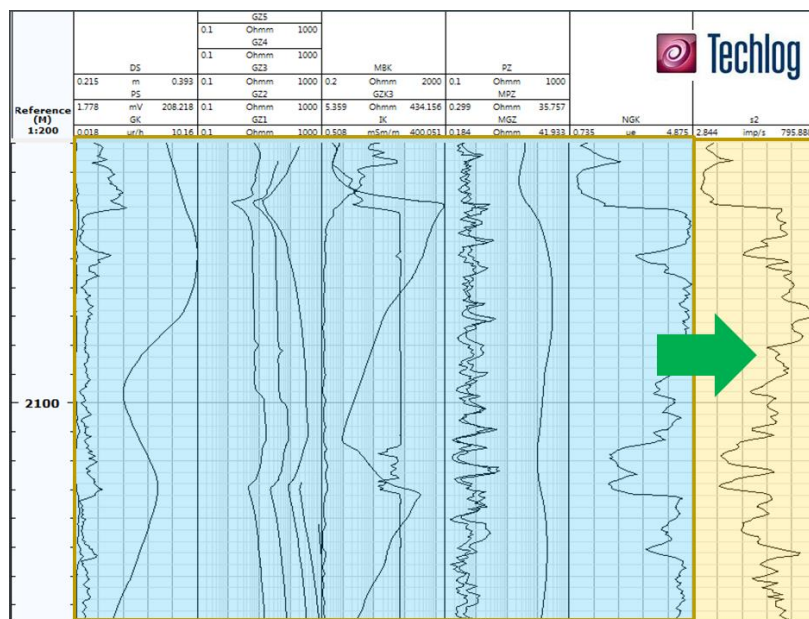


Рисунок 1. Пример набора каротажных кривых

После сбора и обработки данных, формируется обучающая и тестовая выборки для обучения и тестирования алгоритмов соответственно. Распределение тестовых и исследуемых данных составило 20% - 80% соответственно.

Таким образом, у нас есть данные по каротажам и необходимый каротаж, который будет восстанавливаться, это соответствует машинному обучению - обучению с учителем и задаче регрессии. Поэтому в данной работе будет решаться задача регрессии, в общем виде схема решения представлена на рисунке 2.

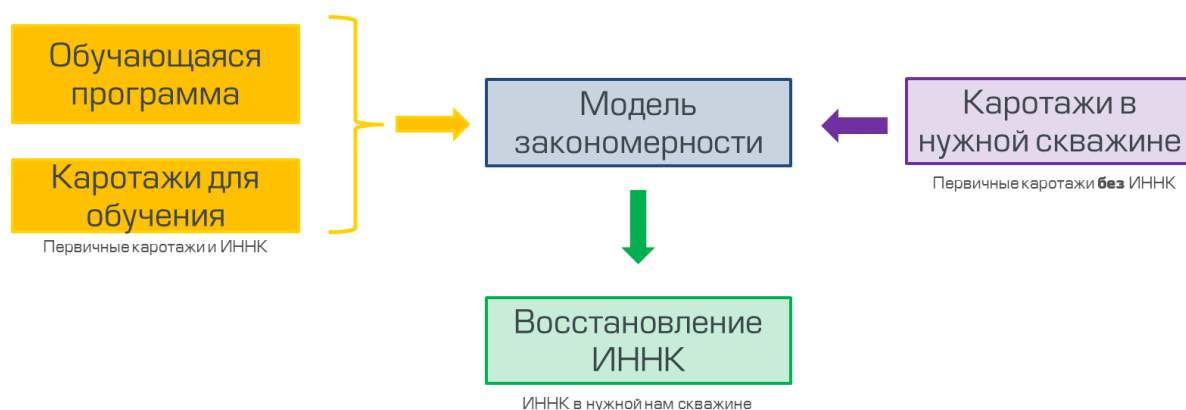


Рисунок 2. Общая схема работы алгоритмов при решении задачи восстановления каротажей

В работе применены 6 алгоритмов решения:

- метод k ближайших соседей
- метод опорных векторов

- градиентный спуск
- случайный лес
- нейронная сеть
- линейная регрессия.

Используя пакет Anaconda и библиотеку SKlearn, были построены модели по всем вышеперечисленным методам.

Таблица 1. Сводная оценка данных алгоритмов

Алгоритм	Среднеквадратическая ошибка	Средняя абсолютная ошибка	R2
kNN	0.007	0.053	0.859
SVM	0.06	0.197	-0.163
SGD	0.03	0.133	0.407
Random Forest	0.006	0.05	0.885
Neural Network	0.013	0.084	0.739
Linear Regression	0.028	0.126	0.459

По результатам тестирования (Таблица 1) видим, что приемлемый результат по R2, среднеквадратичной и абсолютной ошибке показали 3 алгоритма: метод k ближайших соседей, случайный лес и нейросеть.

Далее был проведён визуальный анализ (Рисунок 3).

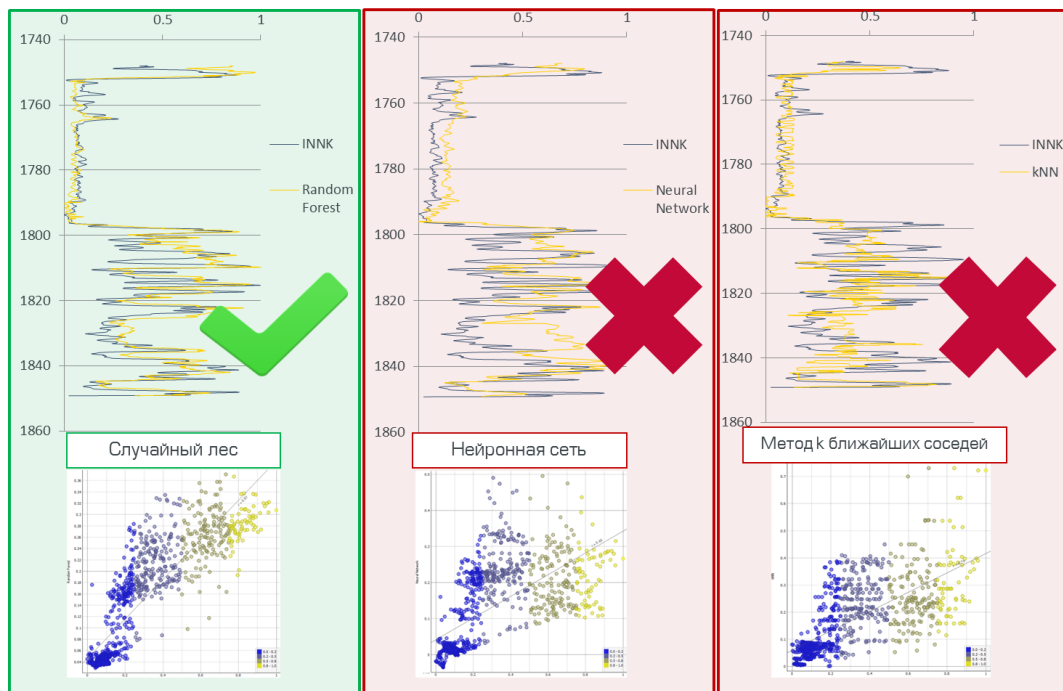


Рисунок 3. Диаграммы рассеяния и визуализация восстановленных кривых в сравнении с фактическими для различных методов

При визуализации данных, мы видим, что наиболее точно кривую ИННК восстановил алгоритм случайного леса с достоверностью $r \sim 0.9$, что позволяет говорить о том, что кривые насыщения возможно восстанавливать с помощью машинного обучения.

Заключение

В результате работы можно говорить о потенциале автоматизации анализа данных геофизики при помощи машинного обучения, а именно, алгоритма случайного леса. Данный алгоритм показал $R^2 \sim 0.9$ и достоверность при визуальном сравнении с тестовой выборкой. При этом аналогичный результат показало применение нейросети, однако применение нейросети оказалось гораздо более затратным по времени, и медленнее на порядок по сравнению со случайным лесом.

Список использованных источников:

1. Ростовцев, В. С. Системы искусственного интеллекта: курс лекций [Текст] / В. С. Ростовцев. – Киров: Изд-во ВятГУ, 2008. – 345 с.
2. The Neural WareTeam: официальный сайт компании Neural Ware. – 2013 [Электронный ресурс]. – Режим доступа: <http://www.neuralware.com/products.jsp> (дата обращения: 20.06.2020).
3. Neuro Dimension Neuro Solutions Software: официальный сайт компании NeuroDimension. – 2013 [Электронный ресурс]. – Режим доступа: <http://www.neurosolutions.com/> (дата обращения: 20.06.2020).
4. UCI Machine Learning Repository: Housing Data Set [Электронный ресурс]. – Режим доступа: <http://archive.ics.uci.edu/ml/datasets/Housing> (дата обращения: 20.06.2020).